



Les identifiants pour les objets numériques

Emmanuelle Bermès

Département de la bibliothèque numérique



Sommaire

- Notion d'identifiants pérennes
- Fonctionnalités
- Etat de l'art
- Les identifiants ARK à la BnF
- Perspectives



Qu'est-ce qu'un identifiant pérenne ?

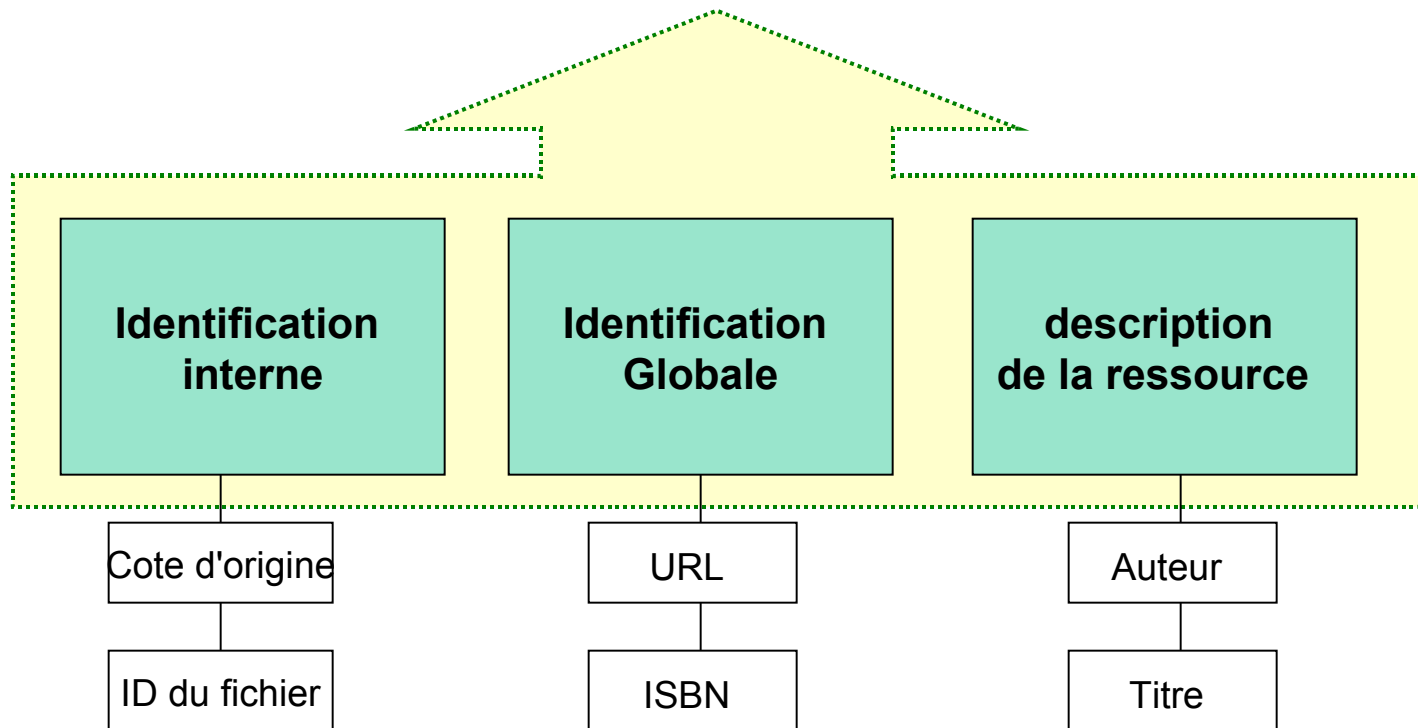
- Une référence stable :
 - nommer la ressource
 - trouver la ressource sur le réseau
 - « citabilité »
- Un identifiant :
 - reconnaître la ressource
 - de manière unique
 - sur le long terme
- Donc deux objectifs :
 - l'accès
 - la préservation



Qu'est-ce qu'un identifiant pérenne ?

Dans le modèle OAIS :

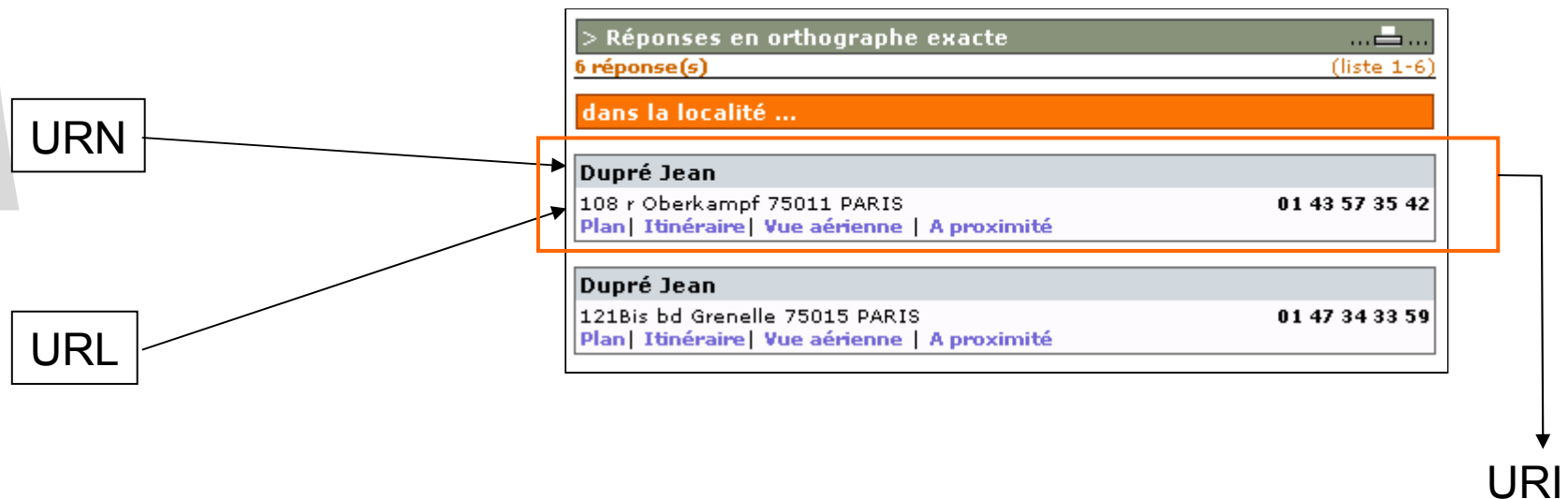
PDI > Information d'identification





Qu'est-ce qu'un identifiant pérenne ?

- On dispose de standards pour identifier les documents sur le Web :
 - URL : *localisation*
 - URN : *nom*
 - URI : *identifiant*
- L'identifiant (URI) peut être le nom (URN) de la ressource ou son adresse / emplacement (URL)
- Aujourd'hui, seules les URL sont interprétées directement par les navigateurs





Qu'est-ce qu'un identifiant pérenne ?

- **Identifiants signifiants**

- · nla.ms-ms51-1 pour la série 1 des Papiers de Sir Edmund Barton;
- · nla.ms-ms51-1-78 pour une lettre de Sir Edmund Barton à Jean M. Barton dans cette série;
- · nla.ms-ms51-1-78-s2 pour la deuxième page de cette lettre;
- · nla.ms-ms51-1-78-s2-a1 pour un détail de cette page.

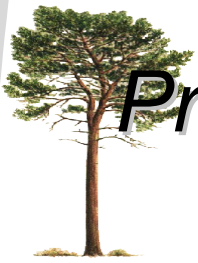
- **Identifiants opaques**

- urn:hdl:**loc.pnp/gsc.5a14460**
- <http://ark.cdlib.org/ark:/13030/kt9x0nc7r1>



Qu'est-ce qu'un identifiant pérenne ?

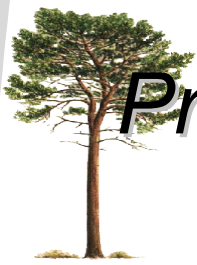
- Identifiants abstraits
 - <http://www.lemonde.fr> : identifiant pérenne du Journal Le Monde mais le contenu change tous les jours
- Identifiants concrets
 - <http://www.lemonde.fr/20070920> : identifiant pérenne du Journal le Monde pour le 20 septembre 2007 (le contenu est stable, la forme peut changer)
 - <http://www.lemonde.fr/20070920/RSS> : identifiant pérenne du Journal le Monde pour le 20 septembre 2007 (le contenu est stable, la forme est stable)



Principales fonctionnalités des identifiants

UNICITE

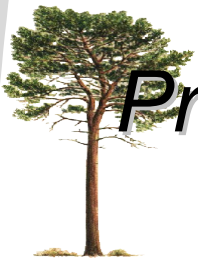
- l'identifiant désigne une ressource et pas une autre
- mais une ressource peut avoir plusieurs identifiants
- Plusieurs niveaux possibles d'unicité
 - unicité locale : au sein d'un système pour un usage interne
 - unicité globale : sur le réseau pour un usage ouvert
- Une même ressource située à plusieurs endroits devrait avoir le même identifiant
 - nécessité d'organisation à l'échelle internationale



Principales fonctionnalités des identifiants

PERENNITE

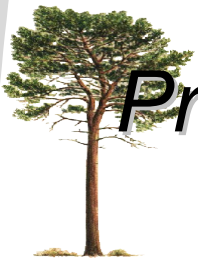
- l'identifiant ne doit pas dépendre de l'emplacement physique de la ressource (URL)
- la pérennité est garantie par l'institution qui attribue les identifiants
 - « autorité nommante »
- L'autorité nommante est identifiée par
 - son nom (BnF)
 - son nom de domaine (bnf.fr) => géré au niveau international
 - un code alphanumérique (123456) => attribué par une autorité internationale



Principales fonctionnalités des identifiants

GRANULARITE

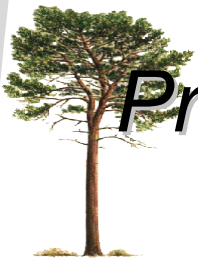
- l'identifiant est applicable à plusieurs niveaux
 - collection
 - document
 - page
 - ...
- Il permet d'accéder à plusieurs versions d'une même ressource
- Il permet d'actionner l'accès à des services ou à des métadonnées
- Attribution d'identifiants indépendants à chaque niveau OU gestion hiérarchique ?



Principales fonctionnalités des identifiants

ADAPTABILITE / EXTENSIBILITE

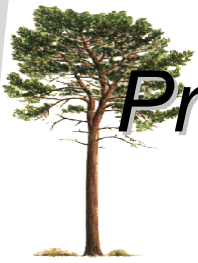
- l'identifiant doit s'adapter à des modèles préexistants si nécessaire
 - ISSN, ISBN
 - cotes
 - précédent système de nommage
- capacité à traiter des identifiants provenant d'un système extérieur
- capacité à s'adapter facilement à de nouveaux types de documents qui n'étaient pas prévus au départ



Principales fonctionnalités des identifiants

VISIBILITE

- l'identifiant peut être utilisé par un navigateur
 - l'outil qu'on utilise pour parcourir le Web
 - interprétation directe (URL)
 - interprétation indirecte (résolveur)
- Fonctions du résolveur
 - interprétation locale des identifiants
 - interprétation globale des identifiants d'un système
 - gestion technique des redirections
 - visibilité (affichage de l'identifiant dans la barre d'adresse)



Principales fonctionnalités des identifiants

INDEPENDANCE

- L'institution contrôle intégralement le système d'identifiants :
 - conventions de nommage
 - architecture logicielle et matérielle
- Eviter ce qui menace la pérennité des identifiants
 - dépendance financière
 - dépendance technique
 - limitations du système par rapport aux fonctionnalités précédemment énoncées



Etat de l'art

- Aujourd'hui il n'existe pas de système qui remplisse parfaitement toutes ces fonctionnalités
- Les points communs :
 - une syntaxe commune : celle des URI
 - `S C H E M E - N A M I N G A U T H O R I T Y - N A M E`
- Les différences :
 - indépendance
 - structure
 - résolveur ...



Handle et DOI

- Une organisation internationale avec un système centralisé
- Caractéristiques techniques
 - Une suite logicielle complète (résolveur, proxy)
 - gestion de métadonnées associées à l'identifiant
 - mais pas de granularité
- Exemples :
 - **hdl:loc.pnp/gsc.5a14460**
 - doi:10.1045/november2005-bartolo
- Un système ancré dans la communauté métier des éditeurs

OAI, PURL et POI



- Le protocole OAI-PMH intègre un système propre d'identifiants
 - oai:bnf.fr:gallica/IFN-6500001
- PURL
 - est maintenu par l'OCLC
 - utilise la transaction HTTP redirect
 - <http://purl.oclc.org/OCLC/OLUC/32821870/2>
 - <http://memory.loc.gov/ammem/browse/ListSome.php?format=Motion+Picture>
- POI
 - est un service qui permet de résoudre directement les identifiants OAI en passant par PURL



URN et INFO

- **urn:** et **info:** sont deux schèmes URI enregistrés auprès de l'IANA
- Leur objectif est de garantir l'unicité d'autres systèmes d'identifiants qui leur sont affiliés mais ne sont pas reconnus comme des URI
 - Ex. : urn:nbn: pour les bibliographies nationales
 - Ex. info:nla: pour les identifiants de la bibliothèque nationale d'Australie
- Ils ne sont pas faits pour être liés à des systèmes de résolution (même si on peut le faire)



- Format d'identification de la California Digital Library (John Kunze)
- Caractéristiques techniques
 - indépendance complète
 - gestion de la visibilité et de la citabilité grâce au concept de « Name Mapping Authority Hostport »
 - gestion répartie des responsabilités
 - gestion de la granularité, des métadonnées, des services grâce au concept de « qualifieur »
- Adopté par le consortium IIPC



Le choix d'un système à la BnF : ARK



Objectif de mise en œuvre à la BnF

- Nos priorités pour un système d'identifiants pérennes fiable et durable :
 - répondre à la fois aux besoins du **système d'archivage** et à ceux de la **communication et de la consultation** des documents, dans le respect de l'existant mais en préfigurant les futures extensions
 - garantir **l'indépendance** de la BnF pour le nommage, s'intégrer à **l'architecture existante**, être un système **ouvert et normalisé** sur le plan international
 - doit fonctionner dans **l'état actuel** des techniques du Web, éviter aux usagers la manipulation d'outils techniques spécifiques et s'intégrer dans leurs pratiques documentaires sur le Web
 - assurer une **meilleure visibilité** et compréhension des contenus pour les **usagers** et pour les robots d'indexation des **moteurs de recherche**



Objectif d'archivage

- Les identifiants ARK permettent de gérer
 - La granularité des paquets d'information
 - Le lien entre les objets d'information et leurs métadonnées
- Couche d'abstraction du stockage : utiliser les identifiants pérennes pour fournir aux applications une vue abstraite sur les données
- Pérennité, extensibilité grâce au système de nommage opaque
- Adaptabilité car il réemploie les anciens systèmes de nommage BnF
- Indépendance (financière, logicielle)
- Mais tout de même une normalisation sur le plan international



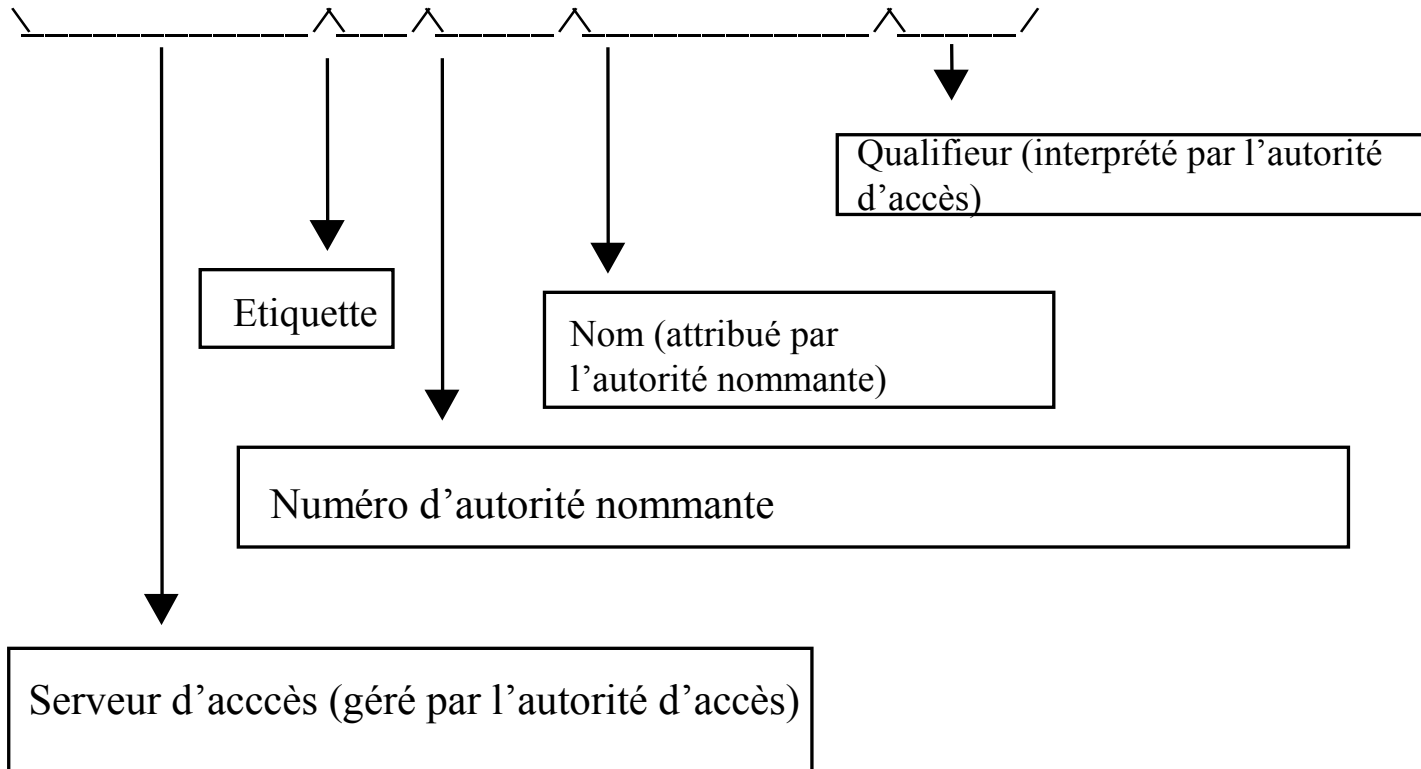
Objectif d'accès

- Les identifiants ARK s'affichent dans la barre d'adresse du navigateur lors de la consultation d'un document numérisé.
- Faciliter des actions simples :
 - ajouter un document dans ses favoris,
 - le citer sur un site Web, dans un mail, sur un blog ou dans un forum
 - citer une page précise dans un livre numérisé, ou référencer une image dans un lot d'images.
- La BnF garantit que cette adresse est pérenne, c'est-à-dire qu'elle ne changera pas, quels que soient les changements d'architecture informatique qui pourraient affecter la bibliothèque numérique.



Le principe du système ARK

`http://ark.bnf.fr/ark:/12148/bpt6k103039f/P122`





Le principe de l'Autorité d'Accès

- C'est un nom de domaine Web
- Il ne fait pas partie de l'identifiant
- Il permet à l'identifiant d'être actionnable.
- A la BnF :
 - <http://catalogue.bnf.fr>
 - pour le contexte catalogue
 - Même comportement par défaut que quand on accède depuis le catalogue
 - <http://gallica.bnf.fr>
 - pour le contexte Gallica
 - Même comportement par défaut que quand on accède depuis Gallica
- Autres : Europeana, plateformes de test, Gallica2...



La structure d'un Nom

- La structure d'un Nom est la suivante :

APPPPNNNNNNNNC

- A : sous-autorité nommante°: une lettre : document numérique (b) ou catalogue (c).
 - PPPP : préfixe (remplace les anciens NUMM, IFN...) opaque pour favoriser l'extensibilité
 - NNNNNNNN est un numéro séquentiel sur 1 à 8 chiffres.
 - C est un caractère de contrôle.
-
- 7,1 e13 identifiants possibles soit l'équivalent 3,5 millions de fois le nombre de documents de la BnF (20 millions).



Les QUALIFIEURS

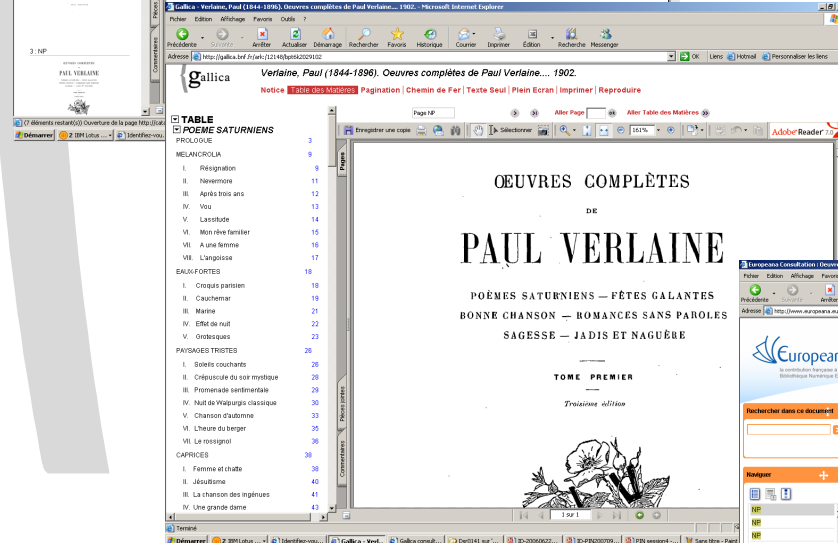
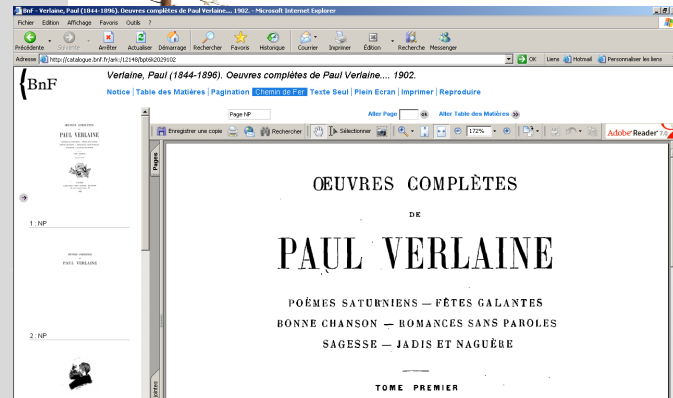
- Deux types de QUALIFIEURS :
 - Hiérarchie
 - Les portions de document : page dans un ouvrage, image dans un lot
 - Service
 - Pour appeler les différentes fonctions du visualiseur : pagination, notice, chemin de fer, etc.



démo

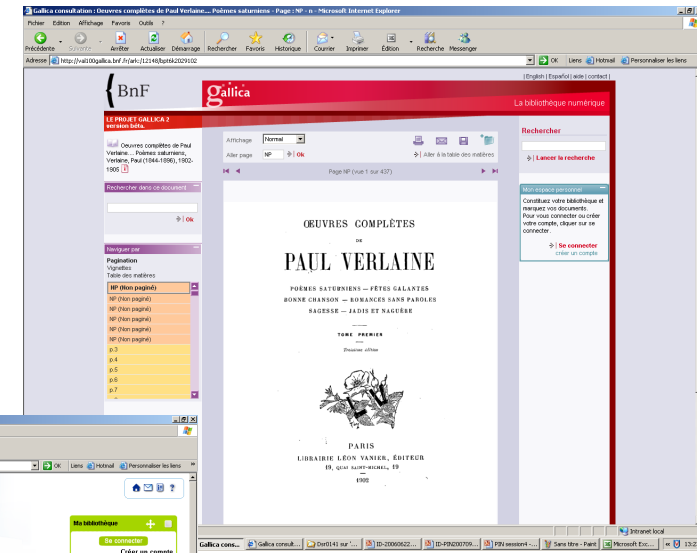
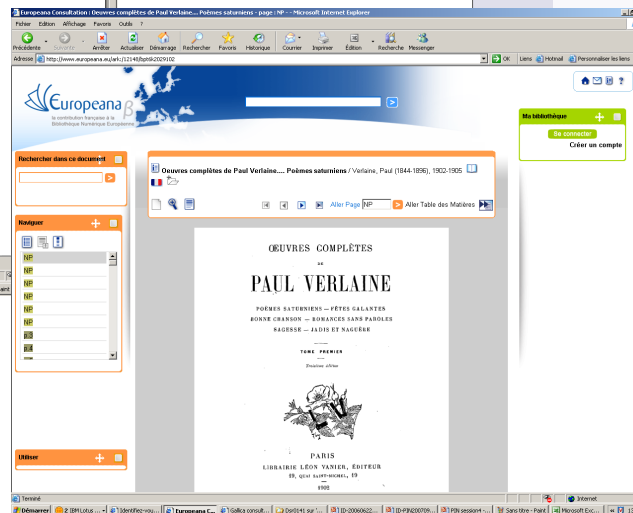
ark:/12148/bpt6k2029102

catalogue



Gallica

Europeana



Gallica2 (tests)



<http://gallica.bnf.fr/ark:/12148/bpt6k2029102/f37.notice>

Gallica - Verlaine, Paul (1844-1896). Oeuvres complètes de Paul Verlaine.... 1902. - Microsoft Internet Explorer

Fichier Edition Affichage Favoris Outils ?

Précédente Suivante Arrêter Actualiser Démarrage Rechercher Favoris Historique Courrier Imprimer Édition Recherche Messenger

Adresse <http://gallica.bnf.fr/ark:/12148/bpt6k2029102/f37.notice> OK Liens Hotmail Personnaliser les liens

gallica Verlaine, Paul (1844-1896). Oeuvres complètes de Paul Verlaine.... 1902.

Notice Table des Matières | Pagination | Chemin de Fer | Texte Seul | Plein Ecran | Imprimer | Reproduire

Page 33 Aller Page Aller Table des Matières

Enregistrer une copie Rechercher Sélectionner 172% Adobe Reader 7.0

Type : texte imprimé, monographie
Auteur(s) : Verlaine, Paul (1844-1896)
Titre(s) : Oeuvres complètes de Paul Verlaine... [Texte imprimé]
Publication : Paris : L. Vanier, 1902-1905
Description matérielle : 5 vol. ; in-8

Comprend : I. Poèmes saturniens. Fêtes galantes. Bonne chanson. [Romances sans paroles.] Sagesse. Jadis et naguère ; IV. Les Poètes maudits. Louise Leclercq. Les Mémoires d'un veuf. Mes hôpitaux. Mes prisons ; V. Confessions. Quinze jours en Hollande. Les Hommes d'aujourd'hui. Portrait

Note(s) : Même contenu que dans les tomes correspondants de la 1re édition conservée sous la cote : [Rés. p-Z-1868. - Contient plusieurs éd.

Notice n° : FRBNF31560807

Cote : NUMM-202910

Pages

Plèces jointes

Commentaires

V

CHANSON D'AUTOMNE

Les sanglots longs
Des violons
De l'automne
Blessent mon cœur
D'une langueur



<http://catalogue.bnf.fr/ark:/12148/cb405059450/PUBLIC>

Catalogue Bn-Opale plus - Notice bibliographique - Microsoft Internet Explorer

Fichier Edition Affichage Favoris Outils ?

Précédente Suivante Arrêter Actualiser Démarrage Rechercher Favoris Historique Courrier Imprimer Edition Recherche Messenger

Adresse http://catalogue.bnf.fr/servlet/NoticeBib;jsessionid=00002EqL8A2_3gbWB7BWT4ZJ6IM:-1?allerA=2&noPage=1&idNoeud=1.3.1.1.1&host=catalogue OK Liens Hotmail Personnaliser les liens

Catalogue Bn-Opale plus Recherche simple > Liste d'entrées titre > Liste des notices > Notice bibliographique

Accueil | Aide

Espace personnel En savoir plus ...

Identifiez vous

Recherche

- simple
- avancée / experte
- équation
- par cote

Recherches spécialisées

Autorités

Rappel de la recherche : TIT : figures

Mes recherches | Mes préférences | Réservations | Panier

rebondir

Affichage public ISBD Intermarc Unimarc

1 2 3 4 5 6

Type : image fixe, monographie

Auteur(s) : Antonini, Annapia (1942-....). Graveur

Titre(s) : Figures [Image fixe] : [estampe] / Annapia Antonini

Publication : Nice : Editions de Francony, 1988

Éditeur : Editions de Francony . Éditeur commercial

Imprimeur / Fabricant : Atelier de taille-douce Tanguy Garric . Imprimeur

Description matérielle : 1 est. : pointe sèche et aquatinte en coul.

Note(s) : Réf. bibl. : Antonini (Annapia).- Catalogue raisonné de l'oeuvre gravé. Paris : 1981, 1987, 1994 (3 vol.), 514

Tirage : 175 ex., XXV, 26 EA

Notice n° : FRBNF40505945

Exemplaire et cote (1)

1 Richelieu - Estampes et photographie - magasin CA20E- 1 (ANTONINI, Annapia) -BOITE FOL support : estampe

ajouter au panier imprimer envoyer par courriel nom@domaine.com ajouter aux favoris



<http://catalogue.bnf.fr/ark:/12148-cb119027577/PUBLIC>

Catalogue Bn-Opale Plus - Notice d'autorité personne - Microsoft Internet Explorer

Fichier Edition Affichage Favoris Outils ?

Précédente Suivante Arrêter Actualiser Démarrage Rechercher Favoris Historique Courrier Imprimer Édition Recherche Messenger

Adresse <http://catalogue.bnf.fr/servlet/autorite;jsessionid=0000tucy4unyYI3c19eAgJ51ful;-1?ID=11902757&idVoeud=1.1&host=catalogue> OK Liens Hotmail Personnaliser les liens

Catalogue Bn-Opale plus Recherche simple > Index auteur > Notice d'autorité personne Accueil Aide

Notice d'autorité personne

Rappel de la recherche : NOM = figuier, louis

Mes recherches | Mes préférences | Réservations | Panier

Espace personnel En savoir plus ...

Identifiez vous

Recherche

- simple
- avancée / experte
- équation
- par cote

Recherches spécialisées

Autorités

Affichage public Intermarc Unimarc

Voir les notices

Figuier, Louis (1819-1894) *forme internationale*

Nationalité(s) : France
Langue(s) : français
Responsabilité(s) exercée(s) sur les documents : Auteur
Naissance : 1819-02-05
Mort : 1894

Savant et littérateur

Forme(s) rejetée(s) :
< Figuier, Louis-Guillaume

Sources : Une histoire du papier peint / écrite par Louis Figuier, [1996]
.- Le Savant du foyer : Louis Figuier et la vulgarisation scientifique au XIXe siècle, 1993. - Larousse 19e s.. - DBF
.- BN Cat. gén. : Figuier (Louis-Guillaume)

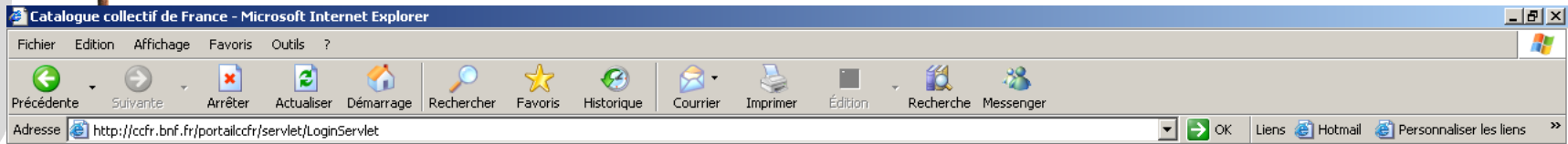
Notice n° : FRBNF11902757 96/09/10

ajouter au panier | imprimer | envoyer par courriel | | ajouter aux favoris

Voir les notices



<http://catalogue.bnf.fr/ark:/12148/cb317290636>



CCFR

ACCUEIL | LOCALISATION | REPERTOIRE | PRET | CONTACTS | ?

Public | ISBD | Professionnel

Accueil
Localisation
Recherche globale
Rech. spécialisées
Répertoire
Historique
Dernière liste

Type document : Livre
Auteur(s) : Arène, Paul (1843-1896), Auteur
Titre : Jean-Des-Figues [Texte imprimé] Paul Arène
Langue : français
Publication : Lyon (rue Président-Carnot, 10) : H. Lardanchet, éditeur, 1921 . -71-Mâcon
Pays : France
Description : 197 p., in-8
Collection : Bibliothèque du bibliophile. Modernes 3
Visualiser la notice dans le catalogue BnF : <http://catalogue.bnf.fr/ark:/12148/cb317290636>

1 - BNF
Bibliothèque nationale de France

Exemplaire(s)

Cote	Bibliothèque	Notes Exemplaire
8-Z-19452 (13) Tolbiac - Rez de jardin - Magasin	Bibliothèque nationale de France	
8-Z-30292 Arsenal - Magasin	Bibliothèque nationale de France	Tirage numéroté. Ex n 982/1000

Catalogue Bn-Opale plus - Notice bibliographique - Microsoft Internet Explorer

Recherche simple > Notice bibliographique

Notice bibliographique

Rappel de la recherche : BNF317290636

Mes recherches | Mes préférences | Réservations | Panier

Affichage public | ISBD | Internarc | Unimarc

Type : texte imprimé, monographie
Auteur(s) : Arène, Paul (1843-1896)
Titre(s) : Jean-Des-Figues [Texte imprimé] / Paul Arène
Publication : Lyon (rue Président-Carnot, 10) : H. Lardanchet, éditeur, 1921 . -71-Mâcon : impr. Protat frères
Description matérielle : 197 p., in-8
Collection : Bibliothèque du bibliophile. Modernes ; 3
Lien à la collection : Bibliothèque du bibliophile. Modernes
Notice n° : FRBNF317290636

Exemplaires et cotes (2)

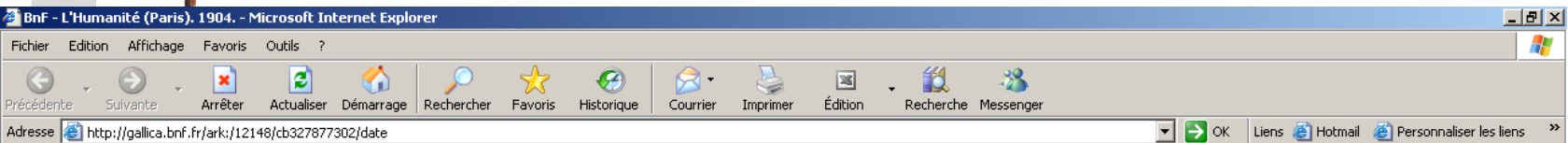
- 1 Tolbiac - Rez-de-jardin - Magasin
8-Z-19452 (13)
support : livre
reproduire
- 2 Arsenal - magasin
8-Z-30292 support : livre
Voir détail exemplaire
reproduire

Consulter | Réserver

ajouter au panier | imprimer | envoyer par courriel | nom@domaine.com | ajouter aux favoris



<http://gallica.bnf.fr/ark:/12148/cb327877302/date>
<http://gallica.bnf.fr/ark:/12148/bpt6k250273r>



L'Humanité (Paris). 1904.

Liste des années

Pour préciser la demande, sélectionner une année de parution :

1910	1911	1912	1913	1904	1905	1906	1907	1908	1909
1920	1921	1922	1923	1924	1925	1926	1927	1928	1929
1930	1931	1932	1933	1934	1935	1936	1937	1938	1939
1940	1941	1942	1943	1944					

Type : texte imprimé, périodique

Titre clé : L'Humanité (Paris)

Titre(s) : L'Humanité [Texte imprimé] : journal socialiste quotidien / dir. politique Jean Jaurès

Mise(s) à jour du titre :

L'Humanité : journal communiste

L'Humanité : organe central du P

Variante(s) historique(s) du titre :

- L'Humanité des débats, 2006 (11

Numerotation : 1904 (8 avr.) ; 1

Publication : Paris : L'Humanité

Description matérielle : Gr. fol.

Note(s) : Du 12 au 27 nov. 1919

en avril 2004. - Les années 1939-1

"L'Humanité : organe central du P

Corrèze)", "Le Trait d'union : solda

Variante(s) de l'adresse : Saint-De

Périodicité : Quotidien

Autre(s) auteur(s) : Parti commu

Jaurès, Jean (1859-1914). Directe

Autre(s) forme(s) du titre :

- Autre forme du titre (figurant sur

Titre(s) en liaison :

- Absorbe : L'Humanité hebdo = IS

- A comme supplément(s) :



gallica

L'Humanité (Paris). 1904.

1904/07/14 (Numéro 88).

Notice | Pagination | Chemin de Fer | Texte Seul | Zoom | Imprimer | Reproduire

Liste des années

1904

14 juillet 1904





Perspectives

Identifiants pérennes dans le système d'archivage



Identifiant "publié" et pour lequel la BnF prend un engagement de pérennité

ark:/12148/bpt6k206840w

Identificateur du Contenu d'Information

Correspondance complexe possible : n-p, mais 1-1 est un cas fréquent

Vue des Entités
« Gestion de données »
et « Accès »

Mise en correspondance
de l'Information de description

bpt6k206840w

Identifiant interne à SPAR

Identificateur d'AIP

Vue de l'Entité
« Stockage »

Adresse

Mise en correspondance de l'Entité "Stockage"

???

Information d'emballage

Contenu d'information

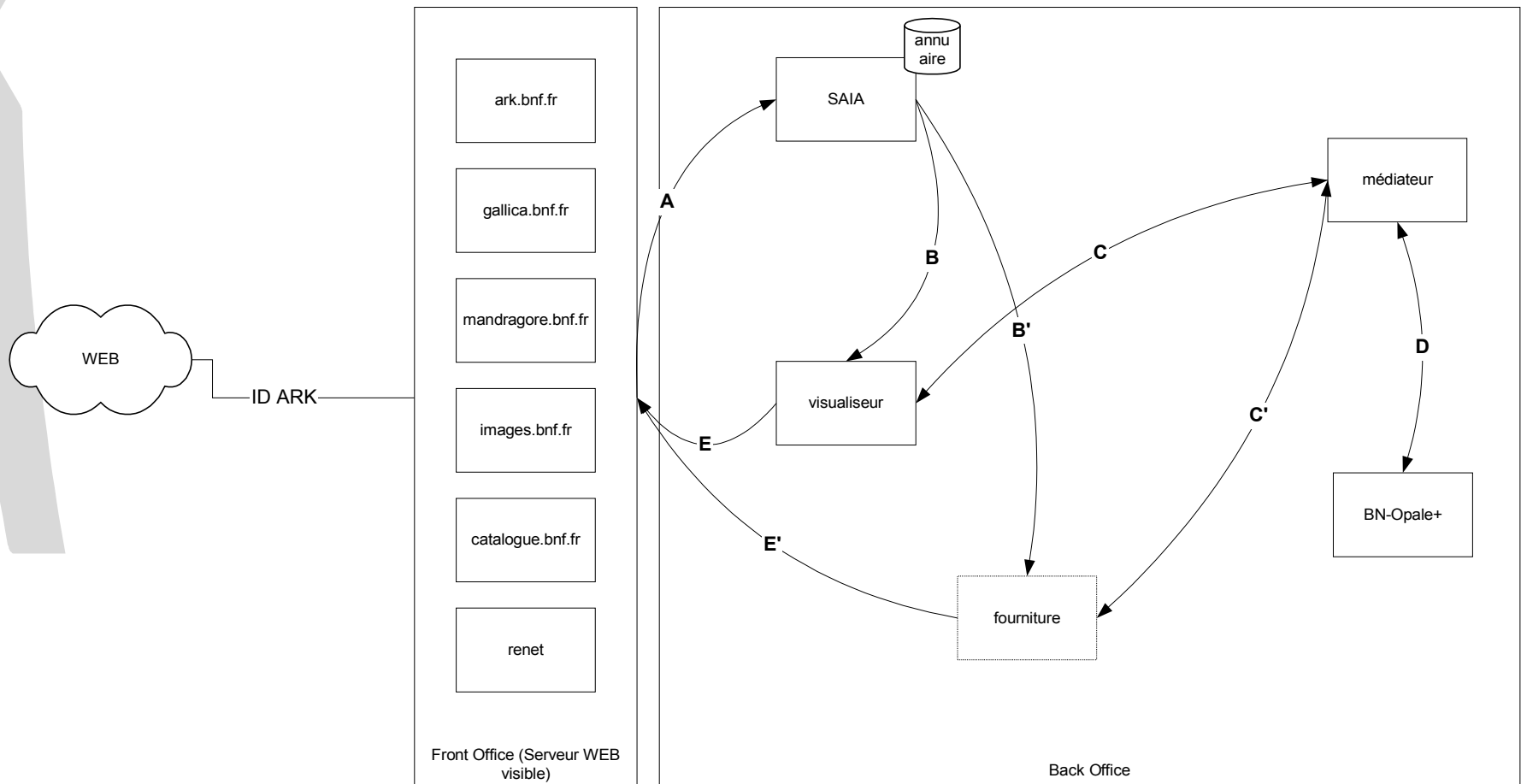
Information de pérennisation

Schéma 5-1 : Vue conceptuelle des relations entre noms et composants d'AIP



Le résolveur (conception)

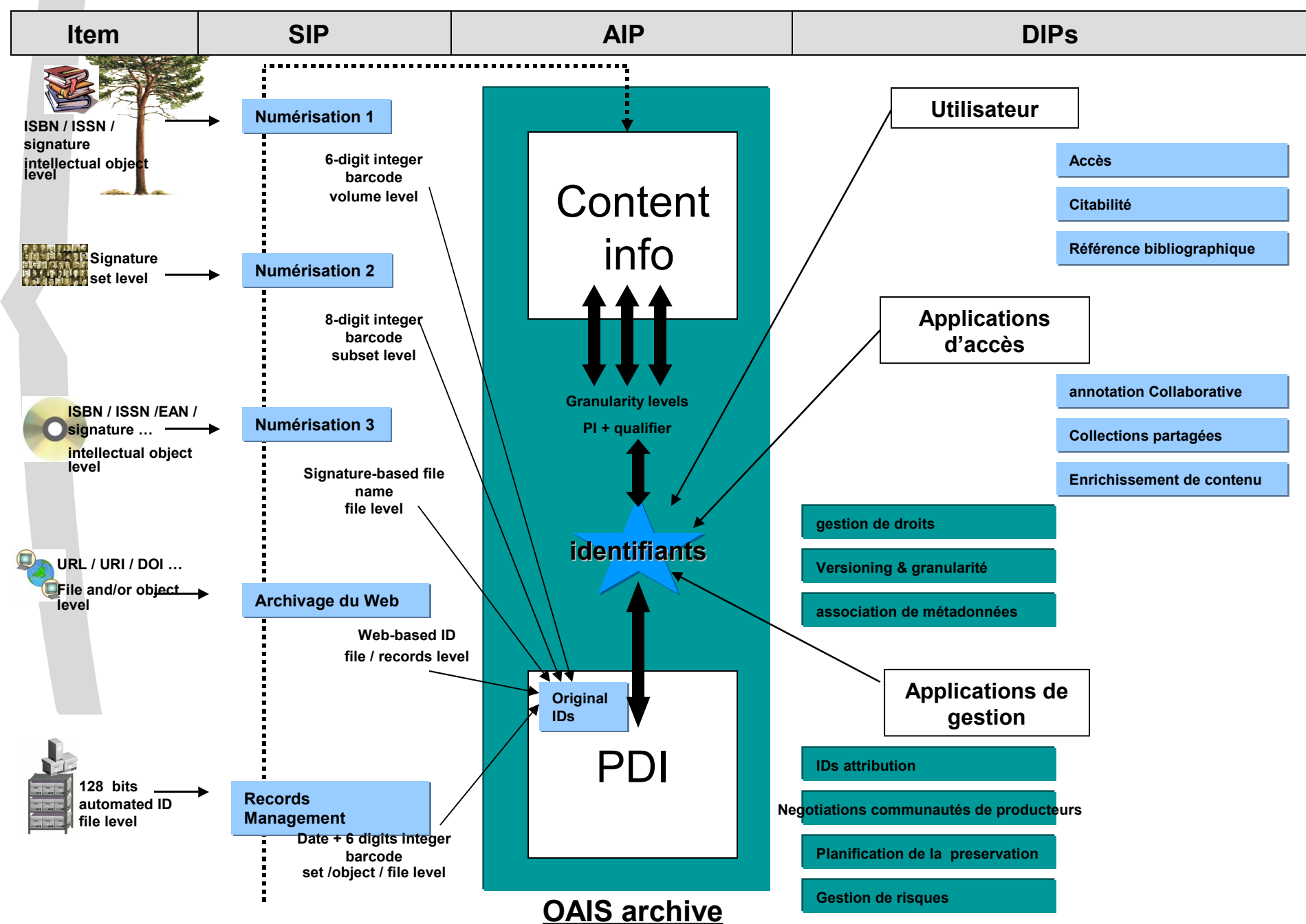
Système d'Automatisation de l'Identification des Archives (SAIA)





Les documents concernés

- Deux types d'objets sont concernés par l'identification pérenne :
 - Les documents numériques versés dans le système d'archivage (DL-WEB => plusieurs millions de documents)
 - Les notices des catalogues (CAT-XML, BN-Opale+) en cours d'instruction
- Autres applications
 - L'ensemble des documents pourraient être traités
 - facilité d'adaptation aux différentes organisations de l'information
 - contrôle de la durée de la vie





Conclusions

- Dans le cadre du futur système de préservation à long terme des documents numériques, les ID ARK représentent :
 - une façon unique de nommer les objets quelle que soit leur nature
 - un point de référence pour gérer les interactions entre les modules du système
 - un moyen d'accès pour améliorer la visibilité des documents sur le Web
 - un élément central de rationalisation et de normalisation du système de gestion des objets numériques